A genomic island linked to ecotype divergence in Atlantic cod

Hansen, Jakob Hemmer; Eg Nielsen, Einar; Therkildsen, Nina O.; Taylor, Martin I.; Ogden, Rob; Geffen, Audrey J.; Bekkevold, Dorte; Helyar, Sarah; Pampoulie, Christophe; Johansen, Torild

Total number of authors:
11

Published in:
Molecular Ecology

Link to article, DOI:
10.1111/mec.12284

Publication date:
2013

A genomic island linked to ecotype divergence in Atlantic cod

Jakob Hemmer-Hansen[1], Einar Eg Nielsen[1], Nina O. Therkildsen[1], Martin I. Taylor[2], Rob Ogden[3], Audrey Geffen[4], Dorte Bekkevold[1], Sarah Helyar[5], Christophe Pampoulie[6], Torild Johansen[7], FishPopTrace Consortium, Gary R. Carvalho[8]

Author affiliations:

1. Section for Marine Living Resources, National Institute of Aquatic Resources, Technical University of Denmark, Vejlsøvej 39, DK-8600 Silkeborg, Denmark
2. School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK
3. TRACE Wildlife Forensics Network, Royal Zoological Society of Scotland, Edinburgh EH12 6TS, UK
4. Department of Biology, University of Bergen, Postboks 7803, N-5020 Bergen, Norway
5. Matís, Vínlandsleið 12, 113 Reykjavík, Iceland
6. Marine Research Institute of Iceland, Skúlagata 4, 121 Reykjavík, Iceland
7. Institute of Marine Research Tromsø, PO Box 6404, N-9294 Tromsø, Norway
8. Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Environment Centre Wales, Bangor University, Bangor, Gwynedd LL57 2UW, UK

Corresponding author:

Jakob Hemmer-Hansen

Technical University of Denmark

National Institute of Aquatic Resources
Vejlsøvej 39
8600 Silkeborg
Denmark
Tel.:+45 35883147
Email: jheha@aqua.dtu.dk

Keywords: Genomic mosaic, ecological divergence, gene flow, evolution, marine fish

Running title: Genomic island of divergence in Atlantic cod

Abstract

The genomic architecture underlying ecological divergence and ecological speciation with gene flow is still largely unknown for most organisms. One central question is whether divergence is genome-wide or localized in "genomic mosaics" during early stages when gene flow is still pronounced. Empirical work has so far been limited, and the relative impacts of gene flow and natural selection on genomic patterns have not been fully explored. Here, we use ecotypes of Atlantic cod to investigate genomic patterns of diversity and population differentiation in a natural system characterized by high gene flow and large effective population sizes, properties which theoretically could restrict divergence in local genomic regions. We identify a genomic region of strong population differentiation, extending over approximately 20 cM, between pairs of migratory and stationary ecotypes examined at two different localities. Furthermore, the region is characterized by markedly reduced levels of genetic diversity in migratory ecotype samples. The results highlight the genomic region, or "genomic island", as potentially associated with ecological divergence and suggest the involvement of a selective sweep. Finally, we also confirm earlier findings of localized genomic differentiation in three other linkage groups associated with divergence among eastern Atlantic populations. Thus, although underlying mechanisms are still unknown, the results suggest that "genomic mosaics" of differentiation may even be found under high levels of gene flow, and that marine fishes may provide insightful model systems for studying and identifying initial targets of selection during ecological divergence.

Introduction

The genomic architecture underlying adaptation to local environments and ultimately ecological speciation (Schluter 2001; Nosil 2012) is poorly understood for most organisms (Wu 2001; Nosil *et al.* 2009; Feder *et al.* 2012). Recent studies have suggested that, during early stages of ecological divergence where gene flow is still on-going, genetic differentiation may be limited to a few specific genomic locations, or "genomic islands", while the majority of the genome remains homogenized by gene flow (Wu 2001; Turner *et al.* 2005; Via & West 2008; Nosil *et al.* 2009; Feder *et al.* 2012a; Feder *et al.* 2012b). Various mechanisms, such as chromosomal inversions (Kirkpatrick & Barton 2006; Feder *et al.* 2011), divergence hitchhiking (Via & West 2008) and processes promoting the genomic co-localization of genes involved in adaptation (Nosil *et al.* 2009; Yeaman & Whitlock 2011), have been proposed as potential mechanisms that would allow differing levels of divergence to evolve within a single genome in the face of gene flow. However, theoretical work has indicated that the conditions, with respect to the relative strengths of selection and gene flow, available for such mechanisms to operate can be relatively restricted (Feder & Nosil 2009; Feder & Nosil 2010; Feder *et al.* 2011; Feder *et al.* 2012b), and that genome-wide divergence should be more common due to effects of reproductive isolation and selection on multiple loci, leading to genome-wide reductions in gene flow (Feder & Nosil 2010). While high gene flow has been predicted to constrain the formation of localized genomic divergence (Feder & Nosil 2009; Feder & Nosil 2010), it has also been suggested that gene flow should promote the clustering of genes involved in local adaptation (Yeaman & Whitlock 2011). Moreover, divergence limited to specific genomic regions should in fact be most readily observable early in the process of divergence, for example between ecotypes (Mallet 2008), rather than at later stages

where gene flow is more restricted and genomic divergence pronounced (Via 2009; Weetman *et al.* 2012).

Hitherto, the investigation of genomic patterns associated with ecological divergence has been restricted to a few, well known model systems, such as walking stick insects (Nosil *et al.* 2008), *Heliconius* butterflies (Nadeau *et al.* 2012), pea aphids (Via & West 2008), malaria mosquitos (Turner *et al.* 2005; Lawniczak *et al.* 2010), coregonid whitefish (Bernatchez *et al.* 2010), three-spined stickleback (Shapiro *et al.* 2004; Colosimo *et al.* 2005; Roesti *et al.* 2012a) and salmonids (Miller *et al.* 2012). Marine fishes provide excellent models for studying interactions between gene flow and selection in the wild since they are often distributed over diverse ecological habitats, and are typically characterized by high levels of gene flow and large effective population sizes (Nielsen *et al.* 2009a). However, although population genetics of non-model organisms, including most marine fishes, has recently moved from analyses of neutral processes towards targeting adaptation to local environments (Luikart *et al.* 2003; Nielsen *et al.* 2009a; Helyar *et al.* 2011), no studies have yet investigated the genomic architecture associated with ecological divergence in these taxa.

Atlantic cod, *Gadus morhua*, has a wide geographical distribution and exploits diverse ecological niches (Mieszkowska *et al.* 2009), ranging from brackish to highly saline environments, and from low temperatures in the Arctic to high and variable temperatures in the southern parts of the distribution (Righton *et al.* 2010). As typical for marine fishes, population structuring is generally shallow (Nielsen *et al.* 2003; O'Leary *et al.* 2007), suggesting high levels of gene flow (Waples 1998) and large effective population sizes (Poulsen *et al.* 2006; Therkildsen *et al.* 2010). Thus, both gene flow and natural selection are predicted to shape genomic patterns of divergence among populations.

Ecologically distinct ecotypes, usually characterised as “migratory” and “stationary” behavioural types, have been described for cod in both eastern and western parts of the Atlantic (Palsson & Thorsteinsson 2003; Robichaud & Rose 2004; Grabowski *et al.* 2011; Nordeide *et al.* 2011). In the eastern Atlantic, these ecotypes are well described in both Iceland and Norway. Migratory individuals are also named “frontal cod” in Iceland and “Northeast Arctic cod” in Norway, while stationary individuals are known also as “coastal cod” in Iceland and “Norwegian coastal cod” in Norway. In general, migratory ecotypes exploit deeper and more offshore habitats at some times of the year compared to stationary individuals which frequent coastal water habitats during their entire life (Palsson & Thorsteinsson 2003; Nordeide *et al.* 2011). Migratory individuals from both locations may also undertake pronounced vertical migrations and cross thermal fronts, formed where warm Atlantic and cold Arctic water meet, during the feeding season (Stensholt 2001; Palsson & Thorsteinsson 2003; Pampoulie *et al.* 2008). Furthermore, Norwegian migratory individuals are characterized by long-distance migrations, for example the ~800 km migration from Lofoten on the Norwegian coast to the feeding areas in the Barents Sea (Jørgensen *et al.* 2008; Sundby & Nakken 2008). In addition to migratory and feeding characteristics, differences in several other life-history related traits, such as growth rate and age at maturity, and in bioenergetics (Pardoe & Marteinsdottir 2009; Nordeide *et al.* 2011) suggest pronounced ecological differences between the two ecotypes (see Nordeide *et al.* (2011) for a comprehensive review). Thus, it is likely that the two ecotypes represent divergent life-history strategies encompassing several behavioural and physiological characteristics of adaptive importance in both Iceland and Norway. Although the ecotypes are ecologically distinct, there is a potential for hybridization between the two types since spawning areas overlap in some regions (Grabowski *et al.* 2011; Nordeide *et al.* 2011). Individuals displaying an intermediate

type of behaviour have been identified through electronic tagging of fish in the wild (Grabowski *et al.* 2011), suggesting that hybridization may occur in nature, but the degree of interbreeding and level of gene flow between ecotypes is presently unknown. Traditionally, morphological characters, such as ear bone structures (otoliths), and single gene markers, such as the membrane protein gene pantophysin (*Pan* I), have been used to designate individuals as either migratory or stationary (Berg & Albert 2003; Pampoulie *et al.* 2008; Wennevik *et al.* 2008). Recently, population genetic work has provided some molecular evidence for adaptive divergence between the ecotypes from Norway (Moen *et al.* 2008; Nielsen *et al.* 2009b), and the finding of consistent migratory profiles over consecutive years for individual fish has suggested a genetic basis for ecotypic divergence in Iceland (Thorsteinsson *et al.* 2012). Yet, the evolutionary relationship between ecotypes is still largely unknown (Nordeide *et al.* 2011) as is the underlying genomic architecture associated with the observed ecotypic differentiation. Furthermore, despite the ecological similarities described above, the evolutionary relationship between Norwegian and Icelandic populations in these parallel systems has not previously been explored.

Here we investigate genomic signatures associated with ecological divergence in a high gene flow scenario. We use the migratory and stationary ecotypes in Atlantic cod as a model system, and examine single nucleotide polymorphisms (SNPs) in population samples of both ecotypes from the two partially isolated systems in Iceland and Norway, along with reference samples from the major population complexes in the species. Information from the Atlantic cod linkage map and the Atlantic cod genome assembly is used to investigate genomic patterns associated with ecotypic divergence.

Materials and Methods

Sampling

Tissue samples of 31-40 adult individuals were collected from each of seven spawning locations and one feeding ground (Fig. 1 and Table 1). Samples representing stationary ecotypes, named "coastal cod" or "stationary cod" in Iceland and "Norwegian coastal cod" in Norway, and migratory ecotypes, named "frontal cod" or "migratory cod" in Iceland and "Northeast Arctic cod" in Norway, were collected from spawning grounds from Iceland and Norway, and individuals were assigned to ecotype based on sampling location and depth (Iceland) and ear bone (otolith) morphology (Norway, see also Wennevik *et al.* (2008)). In Iceland, samples were collected in inshore waters (depth: 58 m), known to be mainly inhabited by the stationary ecotype, and from a deeper offshore location (depth: 135 m), where the migratory ecotype has been suggested to predominate (Pampoulie *et al.* 2006; Pampoulie *et al.* 2008). In Norway, stationary and migratory ecotypes were collected on spawning grounds near the island of Lofoten on the northern Norwegian coast. Due to overlapping spawning areas between the two ecotypes (Grabowski *et al.* 2011; Nordeide *et al.* 2011) there is a risk of including hybrids and/or misclassified individuals in samples collected from spawning areas. Thus, we included a sample from the extreme northern feeding grounds in the Barents Sea (Fig. 1 and Table 1), which are used only by the migratory ecotype (Nordeide *et al.* 2011) and therefore represents a pure "migratory" ecotype sample. In order to relate findings from the stationary/migratory comparison to neighbouring areas, we also included one sample from the highly divergent Baltic Sea (Nielsen *et al.* 2001) and a sample from the North Sea, representing populations near the southernmost part of the distribution in the eastern Atlantic. Finally, one western Atlantic sample was included as an out-group. Thus, with the reference populations, the sampling scheme targeted the major population complexes in the species (O'Leary *et al.* 2007; Bigg *et al.* 2008). The reference populations in the North

Sea and the Baltic Sea are not known to undertake long-distance migrations. However, to allow a direct comparison between the two ecotypes, we refer only to the "stationary" ecotype where it can potentially interbreed with the "migratory" ecotype.

In order to assess temporal stability of genomic patterns, we also analysed temporally replicated samples collected from migratory and stationary populations from Norwegian spawning grounds (Lofoten) and from reference populations in the North Sea and Baltic Sea (Table 1).

Genotyping and initial data filtering

DNA was recovered from samples using the Omega EZNA Tissue DNA kit (Omega Bio-Tek, USA) and subsequently normalised to 50 ng ul^{-1}. Samples were genotyped for 1536 single nucleotide polymorphisms, most of which were originally developed from EST sequences from western Atlantic cod populations ((Hubert *et al.* 2010), see also Table S1), using Illumina's GoldenGate SAM assay on the Bead Array Reader platform. Data were checked against internal sample independent quality controls, clustered and the resulting genotypes then edited manually using the proprietary GenomeStudio software. A replicate individual was included on all plates to ensure genotype reproducibility. Loci with low signal and/or poor clustering were excluded from the analyses.

Linking to the genome assembly

We used the published linkage map consisting of 1310 SNPs (Borza *et al.* 2010) to infer linkage group and position within linkage group for individual SNPs. In addition, a number of SNPs were anchored to the linkage map by mapping the 120 bp flanking sequence of each SNP, available in public data bases, onto the ATLCOD1A genome assembly (Star *et al.* 2011)

using BLASTN with an e-value threshold of 10^{-10}. While these SNPs could be assigned to linkage groups, their position within linkage groups is unknown. We highlight loci in linkage groups previously found to be targets of selection in Atlantic cod (i.e. loci in linkage groups 2, 7 and 12, see Bradbury *et al.* (2010)) along with loci in linkage group 1, which was found to be highly differentiated between ecotypes in this study (see results). The ATLCOD1A genome assembly was also used to estimate the distance (in base pairs) between adjacent SNPs located within the same scaffolds.

Population genetic analyses

For each analysis, loci fixed in all population samples and loci with more than 15% missing genotypes in any sample were removed. Conformance to Hardy-Weinberg equilibrium was tested for each locus in each sample with the package GENETICS v. 1.3.4 for R (R development core team 2011). In order to exclude loci with consistent HWE departures across samples, we excluded loci deviating at the 5% level of significance in more than half of the eight samples. This filtering should assure that loci deviating due to systematic technical or biological reasons were excluded from the analyses. When examining departures from Hardy-Weinberg equilibrium across loci within each sample, we corrected results for multiple testing by using a false discovery rate (FDR) threshold of 5%. FDR correction was done with the package STATS for R, following (Benjamini & Hochberg 1995).

Individual locus pairwise FST coefficients, following (Weir & Cockerham 1984), were estimated with the R package GENELAND (Guillot *et al.* 2005), and mean and 95% confidence intervals were estimated from 1000 data sets generated by bootstrapping over loci.

Population structuring over all loci was examined through correspondence analysis in the package ADEGENET for R (Jombart 2008), using six axes to describe the relationship

among the seven eastern Atlantic population samples. In addition to the full data set, overall pairwise FST was estimated and correspondence analysis conducted on a data set where highly divergent outlier loci identified through Bayesian regression (see below) had been excluded. Loci in the reduced dataset were presumed to be primarily affected by neutral evolutionary forces, such as gene flow and genetic drift. We also investigated the effects of removing loci with global minor allele frequencies below 10% in both the full and the reduced data set, since correspondence analyses gives higher weight to rare alleles (Jombart *et al.* 2009), potentially biasing these analyses.

Observed levels of heterozygosity within samples were estimated for each locus with the R package GENETICS v. 1.3.4, and the R package ZOO was used to calculate moving averages of single locus estimates with a window size of 10 SNPs along each individual linkage group.

A statistical test for FST outliers was conducted by the Bayesian regression method implemented in BAYESCAN 2.1 (Foll & Gaggiotti 2008). The method uses reversible-jump Markov chain Monte Carlo sampling to estimate posterior odds for a model with selection against a model without selection for individual loci. Prior odds for a model without selection were set to 10:1 and 20 pilot runs of each 5000 samplings were used to adjust acceptance rates and to obtain a prior estimate of mean and variance of parameter distributions. Pilot runs were followed by an additional burn in of 50000 and 5000 samplings with a thinning interval of 10 for the estimation of posterior distributions. The false discovery rate was controlled at 5% with the R function plot_bayescan distributed with the package (available from http://cmpg.unibe.ch/software/bayescan/). Outliers were identified in a dataset excluding the highly divergent western Atlantic sample in order to reduce bias due to hierarchical levels of population structuring (Excoffier *et al.* 2009) and to allow a more detailed investigation of

patterns among eastern Atlantic samples. Loci with minor allele frequencies below 2% across all samples were excluded since loci with low information content may bias computations (Beaumont & Balding 2004). The additional filtering step reduced the number of loci to 975 in this analysis. Since loci with low levels of variation may bias outlier tests due to a depression of global FST (Roesti *et al.* 2012b), we estimated global FST for different minor allele frequency thresholds in the eastern Atlantic data set to examine if the chosen threshold had an effect on global FST. In addition, we conducted the outlier test for a dataset where loci with a minor allele frequency below 10% had been excluded in order to examine if outliers were confirmed at a more stringent threshold.

Results

Data filtering and control

Following genotyping and initial data filtering, 295 individuals and 1282 loci were exported for statistical analyses (Table S1). Data quality among retained loci was generally high, with 95% of loci having an average GenCall (GC) score above 0.61 for called genotypes. Initial blast results identified three pairs of identical loci mapping to the same scaffold and position within scaffold (Table S1). One locus from each pair was removed from further analyses. Ten loci were removed from all analyses due to departures from Hardy-Weinberg equilibrium in more than half of the eight samples. After this filtering, only a few loci (between 0 and 11, see Table S1) deviated significantly in each sample, suggesting conformance to Hardy-Weinberg expectations within each of the sampled populations. Following the removal of loci fixed in all population samples and loci with more than 15% missing genotypes in any sample, 1199 loci remained for further analyses when all eight population samples were used. For analyses focusing on the seven eastern Atlantic samples, similar data filtering resulted in a dataset

consisting of 1164 loci. The lower number resulted from a higher number of monomorphic loci among these samples. In addition, observed levels of heterozygosity (Ho) were similar in the eastern Atlantic and Baltic Sea (range of average Ho: 0.23-0.26), but lower than in the western Atlantic (average Ho: 0.34, Table S1), indicating effects from ascertainment bias (see also Discussion).

Genomic distribution of SNPs

The majority of analysed loci, 983 of 1199, were already placed on the linkage map (Table S1). In addition, we were able to assign linkage groups to another 161 SNPs, although with unknown position within linkage groups, through blasting against the ATLCOD1A genome assembly (Table S1). Among the remaining 55 loci, 32 SNPs did not map to a scaffold while 23 SNPs were found in scaffolds that did not contain mapped SNPs. Thus, these loci could not be assigned to any linkage group. While most loci mapped to a scaffold, 227 SNPs mapped to scaffolds containing just the one SNP. The remaining loci were distributed on 236 scaffolds, with the majority of scaffolds containing only few SNPs (Fig. S1). This distribution illustrates the relatively fragmented nature of the current genome assembly. The distribution of distances between adjacent SNPs within scaffolds was also skewed towards lower values (Fig. S2). Thus, the distance to the previous SNP within the same scaffold was below 50,000 bp for most loci and only few pairwise distances were above 1Mb.

Population genetics

Correspondence analysis showed marked differences between the two ecotypes with migratory and stationary samples forming completely separate clusters, each containing both Icelandic and Norwegian samples, when all markers were included in the analysis (Fig. 2a).

311 In contrast, these samples grouped according to geographic origin when a reduced "neutral"
312 data set (i.e. where 87 significant and highly divergent outlier loci had been removed, see also
313 below) was analysed (Fig. 2b). The North Sea and Baltic Sea samples, representing
314 geographically isolated samples, were also genetically isolated in both data sets (Fig. 2).
315 These results were confirmed when loci with a minor allele frequency below 10% were
316 removed (Figure S3), illustrating that these global patterns were robust to the inclusion of rare
317 alleles. The patterns were supported by estimates of pairwise FST (Table S2). With the
318 reduced (neutral) data set, confidence intervals overlapped with zero when comparing
319 ecotypes from spawning grounds within localities. In contrast, although pairwise FST
320 estimates were low, confidence intervals did not overlap with zero when similar ecotypes
321 were compared across the two localities (Table S2).
322 Levels of population differentiation, assessed through individual locus pairwise FST,
323 varied along the linkage groups (Fig. 3; see also Fig. S4 for all comparisons). The pairwise
324 comparisons of migratory and stationary ecotypes collected in both Norway and Iceland (Fig.
325 3a-c) showed markedly increased levels of differentiation for loci in linkage groups 1, 2 and 7
326 in addition to a few loci that were not mapped to a linkage group. In contrast, the pairwise
327 comparisons between similar ecotypes across geographic locations (Fig. 3d and 3e) showed
328 that differentiation was very shallow across all linkage groups. The pairwise comparison
329 between the southernmost eastern Atlantic location from the North Sea and the Norwegian
330 stationary ecotype collected in the northern Atlantic (Fig. 3f) revealed elevated levels of
331 structure for loci in linkage groups 2, 7 and 12, while most remaining loci were weakly
332 differentiated, thus confirming earlier findings of high differentiation in these linkage groups
333 (Bradbury *et al.* 2010). The comparison between the North Sea and the Baltic Sea samples
334 (Fig. 3g), representing reproductively isolated populations (Nielsen *et al.* 2003, see also

335 Discussion), showed elevated differentiation for loci across most linkage groups, as did the
336 comparison between the North Sea and the western Atlantic sample (Fig. 3h).
337 Observed levels of heterozygosity also varied among linkage groups (Fig. 4).
338 Remarkably different patterns in the distribution of heterozygosity were observed among the
339 populations, with dramatic reductions in linkage group 1 in the migratory ecotype samples
340 (Fig. 4a-c). In addition, reduced levels of heterozygosity were observed in linkage group 7 for
341 the migratory ecotype samples (Fig. 4a-c), the North Sea population sample (Fig. 4d) and the
342 western Atlantic sample (Fig. 4h), while the stationary ecotype samples showed increased
343 levels of heterozygosity for the same genomic region (Fig. 4e and 4f).
344 Eighty-seven high FST outlier loci were identified through Bayesian regression on a
345 data set excluding the highly divergent western Atlantic sample and loci with a minor allele
346 frequency below 2%. These outlier loci were primarily located in linkage groups 1, 2, 7 and
347 12 (71 of 87 outliers; Table S3). Global FST changed only slightly (from 0.056 to 0.065)
348 between minor allele frequency thresholds of 0% and 20% (Fig. S5). Changes in global FST
349 were larger for thresholds above 20%, but these analyses only included few loci since most of
350 the loci were removed from analysis at these very high thresholds. In addition, an outlier test
351 including only loci with minor allele frequencies above 10% identified almost the same set of
352 outliers as the test applied on loci with minor allele frequencies above 2% (only four outlier
353 loci were not identified with a threshold of 10%, see Table S3). Thus, results from the outlier
354 test appear very robust to the effects of loci with low information content (see also discussion
355 in Roesti *et al.* (2012b)).
356 Patterns of single locus population differentiation and genetic diversity were
357 confirmed when temporal replicates of the samples from the North Sea, the Baltic sea and
358 both migratory and stationary ecotypes from Norwegian spawning grounds were analysed

359 (Fig. S6 and Fig. S7). Differentiation was increased in linkage groups 1, 2 and 7 in the
360 comparison between the two ecotypes, while differentiation was increased in linkage groups
361 2, 7 and 12 in the comparison between the North Sea and the stationary samples.
362 Differentiation was low across the remaining linkage groups in these comparisons, while
363 differentiation was high across all linkage groups in comparisons involving the Baltic Sea
364 sample (Fig. S6). Genetic diversity was drastically reduced in linkage group 1 in the
365 migratory sample. In addition linkage group 7 showed decreased diversity in the migratory
366 and North Sea samples, while it showed increased diversity in the stationary sample. Finally,
367 loci in linkage group 12 showed decreased diversity in the North Sea sample (Fig. S7). These
368 results indicate temporal stability of observed patterns.

369 A detailed investigation of the loci in linkage group 1 revealed that loci displaying
370 elevated levels of population differentiation between migratory and stationary ecotypes were
371 located between 14.3 and 37.2 cM (Fig. 5 and Table S4). This pattern was evident for both
372 Norwegian and Icelandic comparisons. The previously intensely studied locus in the gene
373 pantophysin (*Pan* I) is located at position 25.1 cM in this linkage group ((Borza *et al.* 2010)
374 and Table S1).

375

376 **Discussion**

377 In addition to identifying a region of high differentiation between ecotypes in linkage group 1,
378 we confirmed earlier findings suggesting selection in linkage groups 2, 7 and 12 in Atlantic
379 cod (Bradbury *et al.* 2010). However, these signals were not specifically associated with the
380 migratory ecotype as was the case for the highly differentiated region in linkage group 1. The
381 region of elevated differentiation between ecotypes extends over 20 cM in a genome subject
382 to high levels of gene flow (see below). Thus, our results suggest that extensive divergence of

local genomic regions may be possible even in situations with extensive gene flow (Yeaman & Whitlock 2011; Weetman *et al.* 2012). In addition, genomic studies of high gene flow scenarios, like ecotypes in marine organisms, may indeed provide valuable model systems for elucidating evolutionary processes at the genomic level associated with ecological divergence (Via 2009; Via 2012).

Origin of migratory ecotype

Despite decades of research on the ecotypes in both Norway and Iceland (Palsson & Thorsteinsson 2003; Nordeide *et al.* 2011), no study has so far directly compared populations from the two regions through the use of a large number of genetic markers. Genetic differentiation between Norway and Iceland (across ecotypes) revealed with neutral genetic markers (Fig 2b and Table S2) suggest reproductive isolation between these locations. Yet, results illustrate marked similarities in genomic signatures associated with ecotypic divergence. Thus, although the description of the ecotypes (or behaviour types) in Icelandic waters has so far only been based on information from data storage tags (Palsson & Thorsteinsson 2003; Pampoulie *et al.* 2008; Grabowski *et al.* 2011), our study confirms the presence of two divergent groups in coastal and deep off-shore locations, respectively.

The region of increased differentiation between ecotypes is also characterized by dramatically reduced levels of diversity in samples representing the migratory ecotype, a classical signal of a selective sweep (Storz 2005). This suggests that initially these populations may have experienced a selective sweep involving the specific region on linkage group 1.

Extremely shallow population differentiation across most of the genome (Fig. 3a-c) as well as the close relationship among populations within geographic locations (across

ecotypes), as estimated with neutral genetic markers (Fig. 2b), suggest two possible scenarios for the origin of migratory ecotype populations. In one scenario, the migratory ecotype arose twice through convergent evolution in two parallel systems (Iceland and Norway) following colonization after the last glacial maximum (LGM) around 21,000 years ago. Similarities within geographic regions (Fig. 2b) could then reflect shared ancestry and recent divergence (Pogson *et al.* 2001) rather than effects from gene flow between ecotypes. However, highly divergent allele lineages for one gene in the region affected by the selective sweep, pantophysin (Pogson & Mesa 2004), suggest that the split of the two ecotypes is ancient compared to the LGM. If the pantophysin gene is representative for the region, these data suggest that recent convergent adaptation is not likely. In contrast, a more parsimonious scenario is that the two ecotypes were already present when deglaciated regions around Iceland and Norway were colonized following the LGM (Kettle *et al.* 2011) and that the geographically based structure at neutral markers is caused by on-going gene flow between ecotypes within localities. This scenario is also consistent with the hypothesized, though still highly speculative, existence of both coastal and off-shore refugia for Atlantic cod during the LGM (Pampoulie *et al.* 2008; Kettle *et al.* 2011). Modelling work has suggested that periods of allopatry, for instance in isolated glacial refugia, could favour the establishment of local genomic differentiation under some models of adaptive divergence (Feder *et al.* 2011). With the current data set it is not possible to determine if secondary contact between ecotypes occurred before or after colonization. However, the combination of highly divergent allele lineages within and extremely shallow differentiation outside the region on linkage group 1 is difficult to explain without a significant role for gene flow. Indeed, if the split is very old and gene flow is not occurring between ecotypes, we would expect to see similar patterns of structuring for neutral markers as those observed for the loci within this specific genomic

region since neutral markers would then reveal common ancestry of ecotypes across locations. In addition, on-going gene flow is also indirectly supported by observations of individuals expressing an intermediate type of behaviour in nature (Grabowski *et al.* 2011), which could suggest on-going hybridization between the ecotypes.

Neutral genetic differentiation between Norway and Iceland (for both ecotypes) also suggests at least partial isolation of the two geographical systems (Waples & Gaggiotti 2006), and that gene flow mostly occurs between ecotypes within the two regions. This gene flow would then be counteracted by on-going selection in the two parallel systems in the specific genomic region in linkage group 1.

Underlying mechanism for genomic differentiation

A number of mechanisms could be responsible for generating and maintaining strong differentiation between ecotypes in the specific region in linkage group 1. If, as suggested above, natural selection is involved, both exogenous (e.g. adaptation to local environmental conditions) and endogenous (i.e. intrinsic incompatibilities) factors could be important and it may be very difficult to disentangle such effects (Bierne *et al.* 2011). While an intrinsic incompatibility unrelated to known ecological and environmental differences cannot be ruled out, the data are also consistent with the alternative interpretation that the migratory ecotype was affected by a selective sweep linked to the unique life-history characteristics known for these populations. It is plausible that the life-history strategy of the migratory ecotype is linked to utilizing high productivity frontal niches in the Arctic for feeding (Stensholt 2001; Grabowski *et al.* 2011), and that the well-described migratory and behavioural characteristics reflect this adaptation. Alternative and more specialized adaptations to different temperature

454 conditions (Righton *et al.* 2010; Grabowski *et al.* 2011) are also likely linked to these
455 differences in life-history strategies between ecotypes.

456 Many studies have discussed selection on the pantophysin gene (e.g. (Pogson 2001;
457 Karlsson & Mork 2003; Case *et al.* 2005; Skarstein *et al.* 2007)), while some authors have
458 noted that observed patterns of linkage disequilibrium within the gene could indicate that
459 selection is instead targeting a linked gene (Fevolden & Pogson 1997). The latter hypothesis
460 is supported by the present study, which suggests that pantophysin may be linked to a large
461 genomic region, potentially harbouring hundreds of genes, rather than the actual target of
462 selection.

463 Although the link between ecotypes and genomic patterns are consistent with patterns
464 resulting from natural selection (through exogenous or endogenous factors) in local
465 populations, alternative explanations could, in principle, also explain our findings. For
466 instance, it has been suggested that transient phases during the fixation process of a globally
467 favourable mutation could generate signals similar to selective sweeps in local populations
468 (Bierne 2010). However, in a scenario of a globally favourable mutation, sweep signals of
469 different magnitudes should be observed in all populations and should be unrelated to specific
470 ecological characteristics (see also (Roesti *et al.* 2012a)). Thus, expected patterns under a
471 globally favourable mutation model are difficult to reconcile with observed patterns, where
472 sweep signals are specifically observed in populations characterized by the migratory life-
473 history strategy. Similarly, structural chromosomal features, such as chromosome
474 centromeres, could potentially explain localized genomic increases in population
475 differentiation due to reduced recombination rates in these regions (Lawniczak *et al.* 2010;
476 Roesti *et al.* 2012a). However, while recombination rate variation would be expected to result
477 in increased levels of differentiation in some parts of the genome, it cannot explain the

478 extreme reduction in diversity observed only in the migratory population samples. Thus, the
479 most plausible explanation remains a balance between local selection and gene flow. Finally,
480 ascertainment bias could have affected some of the analyses conducted in this study since
481 markers were primarily developed from western Atlantic cod populations. Previous studies
482 have not found markedly different levels of diversity in eastern and western Atlantic cod
483 populations (O'Leary *et al.* 2007; Bigg *et al.* 2008), and the lower levels of variation observed
484 in the eastern Atlantic in this study could therefore suggest an effect from ascertainment bias.
485 However, we still do not expect these effects to severely bias the major conclusions drawn
486 from analyses focusing on eastern Atlantic populations, since levels of variation are similar in
487 the eastern Atlantic samples (Table S1) and since all samples in the eastern Atlantic
488 (migratory and stationary populations, in particular) are weakly differentiated from each other
489 and show common divergence from the western Atlantic (Table S2, see also e.g. Rosenblum
490 & Novembre (2007)). Thus, ascertainment bias would be expected to affect eastern Atlantic
491 samples to the same degree.

492 While data suggest increased differentiation over one large genomic region, the
493 relatively modest genome coverage in this study and the fragmented nature of the current cod
494 genome assembly (see Fig. S1 and Fig. S2) does not allow a formal assessment of whether the
495 signals reflect few or several targets of selection (see discussion in Via (2012)). It is possible
496 that future studies applying higher genome coverage may identify more complex patterns of
497 differentiation between cod ecotypes, such as observed in malaria mosquitos (Lawniczak *et*
498 *al.* 2010; Neafsey *et al.* 2010). Similarly, the data do not allow for an assessment of whether
499 divergence hitchhiking, chromosomal rearrangement, such as inversions, or another
500 mechanism is most likely responsible for the observed patterns. It is likely, however, that
501 dense sequencing of the region could elucidate the underlying processes responsible.

Genomic mosaic of differentiation in Atlantic cod

In contrast to patterns observed in linkage group 1, regions of increased differentiation in linkage groups 2, 7 and 12 are not associated with the migratory ecotype samples. These patterns have previously been attributed to coevolution of several genes in response to common environmental conditions (temperature; (Bradbury *et al.* 2010)), but they have not been related to the extremely low levels of differentiation across other parts of the genome, as observed here.

Collectively our results suggest that, on a genome-wide scale, relatively few and potentially large regions, or "genomic islands", could be affected by selection in populations still influenced by gene flow. These patterns are consistent with a "genomic mosaic of divergence" (Wu 2001; Via & West 2008), originally proposed to underlie early stages of ecological divergence in malaria mosquitos and pea aphids (Turner *et al.* 2005; Via & West 2008; Via 2009; White *et al.* 2010; Via 2012). Since these original studies, theoretical and conceptual work has considered whether divergence should be localized or genome-wide during different stages of the "divergence-with-gene-flow continuum" (Feder *et al.* 2012a; Feder *et al.* 2012b; Via 2012). Although the number of empirical studies is increasing, relatively few model systems have so far been studied. While some studies have identified genome-wide patterns of divergence, for instance in walking stick insects (Nosil *et al.* 2008) and three-spined stickleback (Roesti *et al.* 2012a), others have suggested localized divergence, for example in pea aphids (Via & West 2008; Via *et al.* 2012) and *Heliconius* butterflies (Nadeau *et al.* 2012). Interestingly, results from the original model case introducing the "genomic island" metaphor (Turner *et al.* 2005) have been reinterpreted with the availability of genome-wide data to actually reflect pervasive divergence throughout the

526 genome (Lawniczak *et al.* 2010; Neafsey *et al.* 2010), and even studies on the same species
527 under different settings have arrived at different conclusions (Hohenlohe *et al.* 2012; Roesti *et*
528 *al.* 2012a). Thus, so far empirical work has not identified a universal remnant genomic
529 signature following ecological divergence, and it seems likely that different processes operate
530 on different stages of the continuum from panmixia to complete reproductive isolation (Feder
531 *et al.* 2012a).

532 In Atlantic cod, patterns of genomic differentiation associated with clearly
533 differentiated populations from the Baltic Sea and the western Atlantic were different from
534 those observed between weakly differentiated groups. Among highly divergent populations,
535 population differentiation was found across all linkage groups (Fig. 3 and Fig. S4), suggesting
536 reproductive isolation and reduced gene flow (Nielsen *et al.* 2003; Feder *et al.* 2012a).
537 Divergence between the eastern and western Atlantic is believed to be more than 100,000
538 years old, predating the last glacial maximum (Bigg *et al.* 2008). Thus, it may not be
539 surprising that time has allowed genomic differentiation to develop across the Atlantic. In the
540 case of the Baltic Sea, however, Atlantic cod most likely colonized the region following the
541 last glacial retreat from this area around 8,000 years ago (Nielsen *et al.* 2003; Johannesson &
542 Andre 2006). For Atlantic cod and many other marine species, it is therefore plausible that
543 genomic differentiation arose over a relatively short evolutionary time scale following a
544 colonization process involving adaptation, reproductive isolation and increased levels of
545 genetic drift in the Baltic Sea (Johannesson & Andre 2006). Indeed, several life-history
546 characteristics, such as unique sperm activity and egg buoyancy (Nissling & Westin 1997), as
547 well as pronounced genetic differentiation for both neutral and non-neutral genetic markers
548 (Nielsen *et al.* 2003; Nielsen *et al.* 2009b) of Atlantic cod in the Baltic Sea, suggest
549 significant roles for both neutral and non-neutral evolutionary forces in Baltic Sea

populations. The scenarios represented by the Atlantic cod system may therefore represent different stages on the continuum from panmixia to complete isolation (Feder *et al.* 2012a; Via 2012). Importantly, even though the initial split between ecotypes was not recent *per se*, the scenario may still represent an early stage of divergence, i.e. a stage where populations remain connected through significant levels of gene flow (Via 2009). In contrast, reductions in gene flow between highly differentiated groups illustrate that genome-wide effects from neutral evolutionary forces will make it difficult to detect genomic regions associated with initial stages of divergence if populations are investigated at later stages (Via 2009; Via 2012).

Conclusions

The Atlantic cod ecotypes have contributed novel insights on the possible genomic signatures underlying ecological divergence in a high gene flow species. Even though the responsible mechanism and the nature of targets of selection are still unknown, our findings provide additional insights into the long-standing controversy on the interactions between diversifying selection and homogenizing gene flow (Ehrlich & Raven 1969; Mayr 1969; Lenormand 2002; Garant *et al.* 2007). While predictions on the extent and pattern of adaptive divergence can be tested using comparisons of phenotypic traits across populations, analysis at the genomic level allows for unequivocal identification of the integrated effects of selection and gene flow, as well as indicating genes potentially of major effect. Importantly, the frequently documented negative correlations between phenotypic differences and gene flow (Rasanen & Hendry 2008) may be underlain by a much more complex genomic mosaic of response even in high gene flow species (see also Nadeau *et al.* (2012)). Thus, the Atlantic cod ecotypes represent an informative model to study evolution in action (Via 2009), particularly in relation

574 to the dramatic environmental changes predicted for Arctic marine environments under future
575 climate change (Solomon *et al.* 2007).
576
577 **Acknowledgements**
578 We are grateful to Thomas Damm Als for discussions improving the manuscript, and to Sean
579 Rogers and five anonymous reviewers who provided very useful comments on earlier
580 versions of the manuscript. The research leading to these results has received funding from
581 the European Community's Seventh Framework Programme (FP7/2007-2013) under grant
582 agreement n° KBBE-212399 (FishPopTrace).
583
584 **References**

585 Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among
586 populations from genome scans. *Molecular Ecology,* **13**, 969-980.

587 Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a practical and
588 powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-*
589 *Methodological,* **57**, 289-300.

590 Berg E, Albert O (2003) Cod in fjords and coastal waters of North Norway: distribution and
591 variation in length and maturity at age. *ICES Journal of Marine Science,* **60**, 787-797.

592 Bernatchez L, Renaut S, Whiteley AR, et al (2010) On the origin of species: insights from the
593 ecological genomics of lake whitefish. *Philosophical Transactions of the Royal Society of*
594 *London Series B-Biological Sciences,* **365**, 1783-1800.

595 Bierne N (2010) The distinctive footprints of local hitchhiking in a varied environment and
596 global hitchhiking in a subdivided population. *Evolution,* **64**, 3254-3272.

597 Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why
598 genome scans may fail to map local adaptation genes. *Molecular Ecology,* **20**, 2044-2072.

599 Bigg GR, Cunningham CW, Ottersen G, Pogson GH, Wadley MR, Williamson P (2008) Ice-
600 age survival of Atlantic cod: agreement between palaeoecology models and genetics.
601 *Proceedings of the Royal Society of London Series B-Biological Sciences,* **275**, 163-172.

602 Borza T, Higgins B, Simpson G, Bowman S (2010) Integrating the markers Pan I and
603 haemoglobin with the genetic linkage map of Atlantic cod (*Gadus morhua*). *BMC Research*
604 *Notes,* **3**, 261.

605 Bradbury IR, Hubert S, Higgins B, et al (2010) Parallel adaptive evolution of Atlantic cod on
606 both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society*
607 *of London Series B-Biological Sciences,* **277**, 3725-3734.

608 Case R, Hutchinson W, Hauser L, Van Oosterhout C, Carvalho G (2005) Macro- and micro-
609 geographic variation in pantophysin (PanI) allele frequencies in NE Atlantic cod *Gadus*
610 *morhua*. *Marine Ecology-Progress Series,* **301**, 267-278.

611 Colosimo P, Hosemann K, Balabhadra S, et al (2005) Widespread parallel evolution in
612 sticklebacks by repeated fixation of ectodysplasin alleles. *Science,* **307**, 1928-1933.

613 Ehrlich PR, Raven PH (1969) Differentiation of populations. *Science,* **165**, 1228-1232.

614 Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically
615 structured population. *Heredity,* **103**, 285-298.

616 Feder JL, Egan SP, Nosil P (2012a) The genomics of speciation-with-gene-flow. *Trends in*
617 *Genetics,* **28**, 342-350.

618 Feder JL, Gejji R, Powell TH, Nosil P (2011) Adaptive chromosomal divergence driven by
619 mixed geographic mode of evolution. *Evolution,* **65**, 2157-2170.

620 Feder JL, Gejji R, Yeaman S, Nosil P (2012b) Establishment of new mutations under
621 divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society of*
622 *London Series B-Biological Sciences,* **367**, 461-474.

623 Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic
624 islands during ecological speciation. *Evolution*, **64**, 1729-1747.

625 Feder JL, Nosil P (2009) Chromosomal inversions and species differences: when are genes
626 affecting adaptive divergence and reproductive isolation expected to reside within inversions?
627 *Evolution,* **63**, 3061-3075.

628 Fevolden SE, Pogson GH (1997) Genetic divergence at the synaptophysin (Syp I) locus
629 among Norwegian coastal and north-east Arctic populations of Atlantic cod. *Journal of Fish*
630 *Biology,* **51**, 895-908.

631 Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for
632 both dominant and codominant markers: a Bayesian perspective. *Genetics,* **180**, 977-993.

633 Garant D, Forde SE, Hendry AP (2007) The multifarious effects of dispersal and gene flow
634 on contemporary adaptation. *Functional Ecology,* **21**, 434-443.

635 Grabowski TB, Thorsteinsson V, McAdam BJ, Marteinsdottir G (2011) Evidence of
636 segregated spawning in a single marine fish stock: Sympatric divergence of ecotypes in
637 Icelandic cod? *PLoS ONE,* **6**, e17528.

638 Guillot G, Mortier F, Estoup A (2005) GENELAND: a computer package for landscape
639 genetics. *Molecular Ecology Notes,* **5**, 712-715.

640 Helyar SJ, Hemmer-Hansen J, Bekkevold D, et al (2011) Application of SNPs for population
641 genetics of nonmodel organisms: new opportunities and challenges. *Molecular Ecology*
642 *Resources,* **11**, 123-136.

643 Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium
644 and parallel adaptive divergence across threespine stickleback genomes. *Philosophical*
645 *Transactions of the Royal Society of London Series B-Biological Sciences,* **367**, 395-408.

646 Hubert S, Higgins B, Borza T, Bowman S (2010) Development of a SNP resource and a
647 genetic linkage map for Atlantic cod (*Gadus morhua*). *BMC Genomics,* **11**, 191.

648 Johannesson K, Andre C (2006) Life on the margin: genetic isolation and diversity loss in a
649 peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology,* **15**, 2013-2029.

650 Jombart T, Pontier D, Dufour AB (2009) Genetic markers in the playground of multivariate
651 analysis. *Heredity,* **102**, 330-341.

652 Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers.
653 *Bioinformatics,* **24**, 1403-1405.

654 Jørgensen C, Dunlop ES, Opdal AF, Fiksen O (2008) The evolution of spawning migrations:
655 State dependence and fishing-induced changes. *Ecology,* **89**, 3436-3448.

656 Karlsson S, Mork J (2003) Selection-induced variation at the pantophysin locus (PanI) in a
657 Norwegian fjord population of cod (*Gadus morhua* L.). *Molecular Ecology,* **12**, 3265-3274.

658 Kettle AJ, Morales-Muniz A, Rosello-Izquierdo E, Heinrich D, Vollestad LA (2011) Refugia
659 of marine fish in the northeast Atlantic during the last glacial maximum: concordant
660 assessment from archaeozoology and palaeotemperature reconstructions. *Climate of the Past,*
661 **7**, 181-201.

662 Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation.
663 *Genetics,* **173**, 419-434.

664 Lawniczak MK, Emrich SJ, Holloway AK, et al (2010) Widespread divergence between
665 incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science,* **330**,
666 512-514.

667 Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology &*
668 *Evolution,* **17**, 183-189.

669 Luikart G, England P, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of
670 population genomics: From genotyping to genome typing. *Nature Reviews Genetics,* **4**, 981-
671 994.

672 Mallet J (2008) Hybridization, ecological races and the nature of species: empirical evidence
673 for the ease of speciation. *Philosophical Transactions of the Royal Society of London Series*
674 *B-Biological Sciences,* **363**, 2971-2986.

675 Mayr E (1969) *Animal Species and Evolution*. Harvard University Press, Cambridge, MA.

676 Mieszkowska N, Genner MJ, Hawkins SJ, Sims DW (2009) Chapter 3. Effects of climate
677 change and commercial fishing on Atlantic cod Gadus morhua. *Advances in Marine Biology,*
678 **56**, 213-273.

679 Miller MR, Brunelli JP, Wheeler PA, et al (2012) A conserved haplotype controls parallel
680 adaptation in geographically distant salmonid populations. *Molecular Ecology,* **21**, 237-249.

681 Moen T, Hayes B, Nilsen F, et al (2008) Identification and characterisation of novel SNP
682 markers in Atlantic cod: evidence for directional selection. *BMC Genetics,* **9**, 18.

683 Nadeau NJ, Whibley A, Jones RT, et al (2012) Genomic islands of divergence in hybridizing
684 *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical*
685 *Transactions of the Royal Society of London Series B-Biological Sciences,* **367**, 343-353.

686 Neafsey DE, Lawniczak MK, Park DJ, et al (2010) SNP genotyping defines complex gene-
687 flow boundaries among African malaria vector mosquitoes. *Science,* **330**, 514-517.

688 Nielsen EE, Hansen MM, Schmidt C, Meldrup D, Grønkjær P (2001) Population of origin of
689 Atlantic cod. *Nature,* **413**, 272.

690 Nielsen E, Hansen M, Ruzzante D, Meldrup D, Gronkjaer P (2003) Evidence of a hybrid-zone
691 in Atlantic cod (*Gadus morhua*) in the Baltic and the Danish Belt Sea revealed by individual
692 admixture analysis. *Molecular Ecology,* **12**, 1497-1508.

693 Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009a) Population genomics of
694 marine fishes: identifying adaptive variation in space and time. *Molecular Ecology,* **18**, 3128-
695 3150.

696 Nielsen EE, Hemmer-Hansen J, Poulsen NA, et al (2009b) Genomic signatures of local
697 directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*).
698 *BMC Evolutionary Biology,* **9**, 276.

699 Nissling A, Westin L (1997) Salinity requirements for successful spawning of Baltic and Belt
700 Sea cod and the potential for cod stock interactions in the Baltic Sea. *Marine Ecology-*
701 *Progress Series,* **152**, 261-271.

702 Nordeide JT, Johansen SD, Jorgensen TE, Karlsen BO, Moum T (2011) Population
703 connectivity among migratory and stationary cod *Gadus morhua* in the Northeast Atlantic-A
704 review of 80 years of study. *Marine Ecology-Progress Series,* **435**, 269-283.

705 Nosil P (2012) *Ecological Speciation*, 1st edn. Oxford University Press, New York.

706 Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-
707 stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution,*
708 **62**, 316-336.

709 Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic
710 divergence. *Molecular Ecology,* **18**, 375-402.

711 O'Leary DB, Coughlan J, Dillane E, McCarthy TV, Cross TF (2007) Microsatellite variation
712 in cod *Gadus morhua* throughout its geographic range. *Journal of Fish Biology,* **70**, 310-335.

713 Palsson O, Thorsteinsson V (2003) Migration patterns, ambient temperature, and growth of
714 Icelandic cod (*Gadus morhua*): evidence from storage tag data. *Canadian Journal of*
715 *Fisheries and Aquatic Sciences,* **60**, 1409-1423.

716 Pampoulie C, Jakobsdottir KB, Marteinsdottir G, Thorsteinsson V (2008) Are vertical
717 behaviour patterns related to the pantophysin locus in the Atlantic cod (*Gadus morhua* L.)?
718 *Behavior Genetics,* **38**, 76-81.

719 Pampoulie C, Ruzzante DE, Chosson V, et al (2006) The genetic structure of Atlantic cod
720 (*Gadus morhua*) around Iceland: insight from microsatellites, the Pan I locus, and tagging
721 experiments. *Canadian Journal of Fisheries and Aquatic Sciences,* **63**, 2660-2674.

722 Pampoulie C, Stefansson MO, Jorundsdottir TD, Danilowicz BS, Danielsdottir AK (2008)
723 Recolonization history and large-scale dispersal in the open sea: the case study of the North
724 Atlantic cod, *Gadus morhua* L. *Biological Journal of the Linnean Society,* **94**, 315-329.

725 Pardoe H, Marteinsdottir G (2009) Contrasting trends in two condition indices: bathymetric
726 and spatial variation in autumn condition of Icelandic cod *Gadus morhua*. *Journal of Fish*
727 *Biology,* **75**, 282-289.

728 Pogson GH (2001) Nucleotide polymorphism and natural selection at the pantophysin (Pan I)
729 locus in the Atlantic cod, *Gadus morhua* (L.). *Genetics,* **157**, 317-330.

730 Pogson GH, Mesa KA (2004) Positive Darwinian selection at the pantophysin (Pan I) locus in
731 marine gadid fishes. *Molecular Biology and Evolution,* **21**, 65-75.

732 Pogson GH, Taggart CT, Mesa KA, Boutilier RG (2001) Isolation by distance in the Atlantic
733 cod, *Gadus morhua*, at large and small geographic scales. *Evolution,* **55**, 131-146.

734 Poulsen N, Nielsen E, Schierup M, Loeschcke V, Gronkjaer P (2006) Long-term stability and
735 effective population size in North Sea and Baltic Sea cod (*Gadus morhua*). *Molecular*
736 *Ecology,* **15**, 321-331.

737 Rasanen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and
738 gene flow when ecology drives diversification. *Ecology Letters,* **11**, 624-636.

739 Righton DA, Andersen KH, Neat F, et al (2010) Thermal niche of Atlantic cod *Gadus*
740 *morhua*: limits, tolerance and optima. *Marine Ecology-Progress Series,* **420**, 1-U344.

741 Robichaud D, Rose G (2004) Migratory behaviour and range in Atlantic cod: inference from a
742 century of tagging. *Fish and Fisheries,* **5**, 185-214.

743 Roesti M, Hendry AP, Salzburger W, Berner D (2012a) Genome divergence during
744 evolutionary diversification as revealed in replicate lake-stream stickleback population pairs.
745 *Molecular Ecology*, **21**, 2852–2862.

746 Roesti M, Salzburger W, Berner D (2012b) Uninformative polymorphisms bias genome scans
747 for signatures of selection. *BMC Evolutionary Biology,* **12**, 94.

748 Rosenblum EB, Novembre J (2007) Ascertainment bias in spatially structured populations: a
749 case study in the eastern fence lizard. *The Journal of Heredity,* **98**, 331-336.

750 Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution,* **16**, 372-
751 380.

752 Shapiro M, Marks M, Peichel C, et al (2004) Genetic and developmental basis of evolutionary
753 pelvic reduction in threespine sticklebacks. *Nature,* **428**, 717-723.

754 Skarstein TH, Westgaard J, Fevolden S (2007) Comparing microsatellite variation in north-
755 east Atlantic cod (*Gadus morhua* L.) to genetic structuring as revealed by the pantophysin
756 (Pan I) locus. *Journal of Fish Biology,* **70**, 271-290.

757 Solomon S, Qin D, Manning M*, et al*, eds (2007) *Climate Change 2007: The Physical Science*
758 *Basis. Contribution of Working Group I to the Fourth Assessment Report of the*
759 *Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge,
760 United Kingdom and New York, NY, USA.

761 Star B, Nederbragt AJ, Jentoft S, et al (2011) The genome sequence of Atlantic cod reveals a
762 unique immune system. *Nature,* **477**, 207-210.

763 Stensholt B (2001) Cod migration patterns in relation to temperature: analysis of storage tag
764 data. *ICES Journal of Marine Science,* **58**, 770-793.

765 Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population
766 divergence. *Molecular Ecology,* **14**, 671-688.

767 Sundby S, Nakken O (2008) Spatial shifts in spawning habitats of Arcto-Norwegian cod
768 related to multidecadal climate oscillations and climate change. *ICES Journal of Marine*
769 *Science,* **65**, 953-962.

770 Therkildsen NO, Nielsen EE, Swain DP, Pedersen JS (2010) Large effective population size
771 and temporal genetic stability in Atlantic cod (*Gadus morhua*) in the southern Gulf of St.
772 Lawrence. *Canadian Journal of Fisheries and Aquatic Sciences,* **67**, 1585-1595.

773 Thorsteinsson V, Palsson OK, Tomasson GG, Jonsdottir IG, Pampoulie C (2012) Consistency
774 in the behaviour types of the Atlantic cod: repeatability, timing of migration and geo-location.
775 *Marine Ecology-Progress Series,* **462**, 251-260.

776 Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles*
777 *gambiae*. *PLoS Biology,* **3**, e285.

778 Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological
779 speciation-with-gene-flow. *Philosophical Transactions of the Royal Society of London Series*
780 *B-Biological Sciences,* **367**, 451-460.

781 Via S (2009) Natural selection in action during speciation. *Proceedings of the National*
782 *Academy of Sciences of the United States of America,* **106 (Suppl 1)**, 9939-9946.

783 Via S, Conte G, Mason-Foley C, Mills K (2012) Localizing F(ST) outliers on a QTL map
784 reveals evidence for large genomic regions of reduced gene exchange during speciation-with-
785 gene-flow. *Molecular Ecology, **21***, 5546–5560.

786 Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological
787 speciation. *Molecular Ecology,* **17**, 4334-4345.

788 Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some
789 genetic methods for identifying the number of gene pools and their degree of connectivity.
790 *Molecular Ecology,* **15**, 1419-1439.

791 Waples R (1998) Separating the wheat from the chaff: Patterns of genetic differentiation in
792 high gene flow species. *Journal of Heredity,* **89**, 438-450.

Weetman D, Wilding CS, Steen K, Pinto J, Donnelly MJ (2012) Gene flow-dependent genomic divergence between *Anopheles gambiae* M and S Forms. *Molecular Biology and Evolution,* **29**, 279-291.

Weir B, Cockerham C (1984) Estimating F-statistics for the analysis of population structure. *Evolution,* **38**, 1358-1370.

Wennevik V, Jorstad KE, Dahle G, Fevolden S (2008) Mixed stock analysis and the power of different classes of molecular markers in discriminating coastal and oceanic Atlantic cod (*Gadus morhua* L.) on the Lofoten spawning grounds, Northern Norway. *Hydrobiologia,* **606**, 7-25.

White BJ, Cheng C, Simard F, Costantini C, Besansky NJ (2010) Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Molecular Ecology,* **19**, 925-939.

Wu C (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology,* **14**, 851-865.

Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution,* **65**, 1897-1911.

Data Accessibility

Novel SNPs analysed in this study are available in GenBank (dbSNP) under accession numbers ss678251294-ss678251301. Individual SNP genotypes have been deposited in the DRYAD data repository (doi:10.5061/dryad.9gf10).

814

815 **Author Contributions Box**

816 JHH and EEN designed the study with input from MIT, RO, DB, SH and GRC. AG, CP and
817 TJ contributed samples. JHH and NOT analysed the data. JHH wrote the paper with
818 contributions from all authors.

819

820 **Figure legends**

821 *Figure 1*

822 Locations of samples included in the present study. See Table 1 for detailed sample
823 information.

824

825 *Figure 2*

826 Population relationships among eastern Atlantic samples based on correspondence analysis
827 with all markers (a, 1164 loci) and with neutral markers only (b, 1077 loci).

828

829 *Figure 3*

830 Estimates of pairwise levels of population differentiation (Weir and Cockerhams θ (Weir &
831 Cockerham 1984)) based on 1199 loci ordered by position within linkage groups between (a)
832 Norway migratory on spawning grounds and Norway stationary, (b) Norway migratory on
833 feeding grounds and Norway stationary, (c) Iceland migratory and Iceland stationary, (d)
834 Norway migratory on spawning grounds and Iceland migratory, (e) Norway stationary and
835 Iceland stationary, (f) Norway stationary and North Sea, (g) North Sea and Baltic Sea and (h)
836 North Sea and western Atlantic. Horizontal dashed and dotted lines represent mean and 95th
837 percentiles generated by bootstrapping over loci. Loci in linkage groups 1, 2, 7 and 12 are

838 coloured red, blue, green and purple, respectively, while additional linkage groups are
839 coloured in alternating shades of grey and loci with unknown linkage group are shown in
840 black. Location within linkage group is unknown for loci to the right of the vertical line.
841
842 *Figure 4*
843 Observed levels of heterozygosity based on 983 loci with known linkage group position,
844 estimated as moving averages within linkage groups, in (a) Norway migratory on spawning
845 grounds, (b) Norway migratory on feeding grounds, (c) Iceland migratory, (d) North Sea, (e)
846 Norway stationary, (f) Iceland stationary, (g) Baltic Sea and (h) western Atlantic. Horizontal
847 dashed line marks the 1st percentile over all linkage groups. Estimates for linkage groups 1, 2,
848 7 and 12 are coloured red, blue, green and purple, respectively, while additional linkage
849 groups are coloured in alternating shades of grey.
850
851 *Figure 5*
852 Pairwise FST, estimated by Weir and Cockerhams θ (Weir & Cockerham 1984), between
853 migratory and stationary ecotypes from Norway (a) and Iceland (b) for 57 loci with known
854 linkage group position in linkage group 1. Loci identified as FST outliers by Bayesian
855 regression are shown in red.
856
857
858
859
860
861

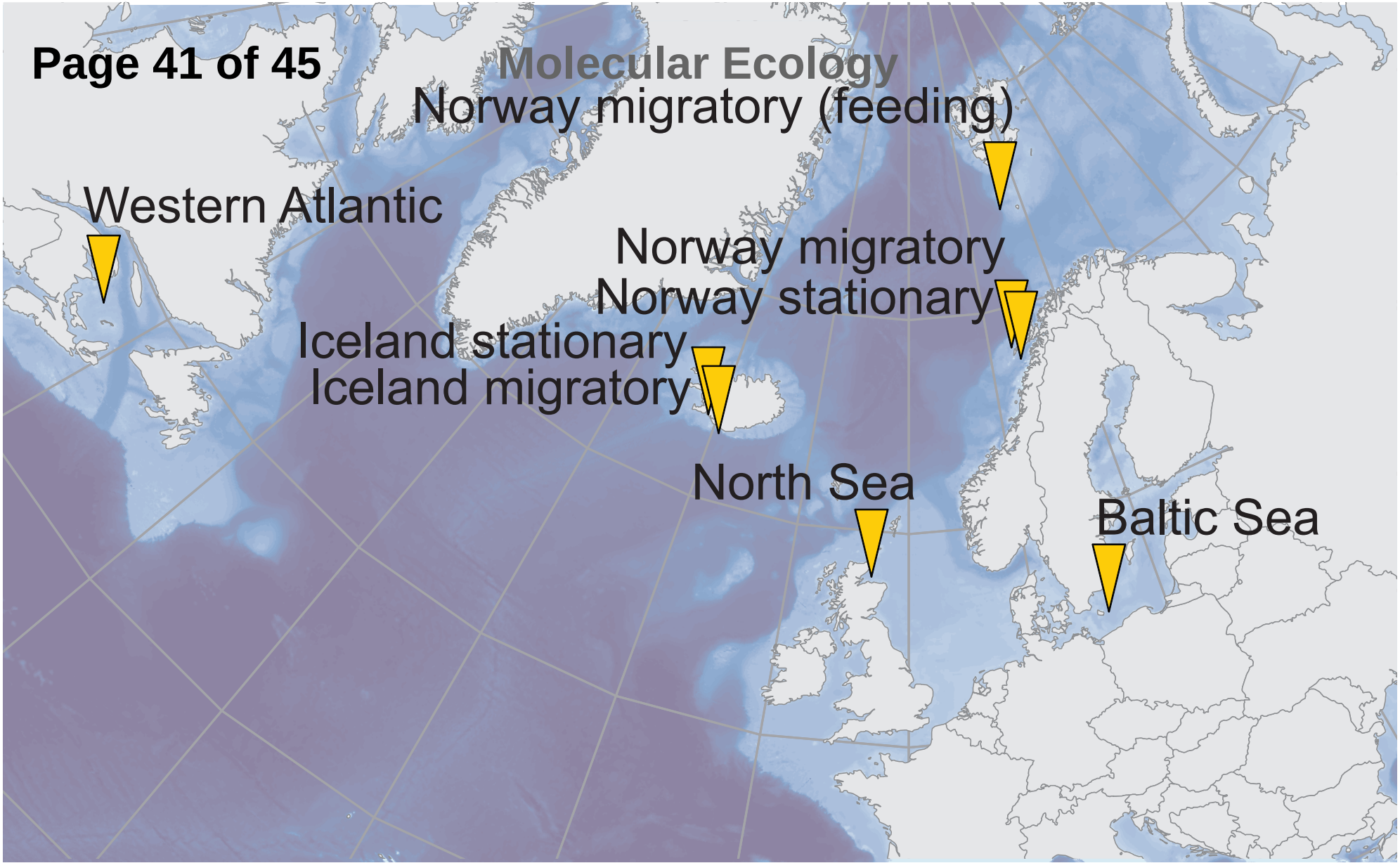
862 **Tables**

863 *Table 1*

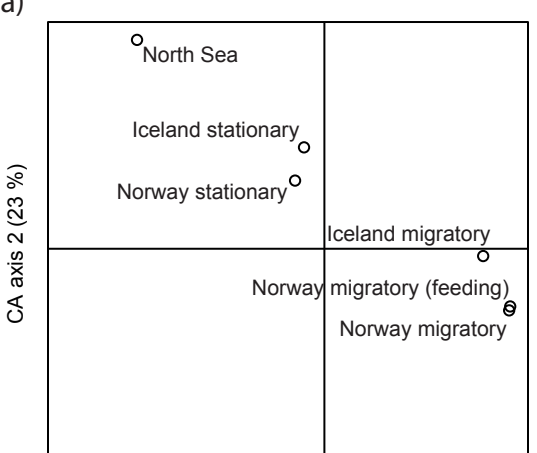
864 Samples of Atlantic cod included in the present study.

Sample	Sample size	Latitude	Longitude	Sampling month/year
Ecotype samples				
Norway migratory (feeding)	35	75.64	16.82	August/2009
Norway migratory (spawning)	35	67.33	11.38	March/2009
Norway stationary	31	68.15	14.48	March/2009
Iceland migratory	39	63.20	-19.30	April/2002
Iceland stationary	38	63.49	-21.05	April/2002
Reference samples				
North Sea	38	56.91	7.83	February/2007
Baltic Sea	40	55.04	15.30	March/2006 and April/2007
Western Atlantic	39	48.01	-63.55	May/2008
Temporal replicates				
Norway migratory	35	68.35	12.14	April/2003
Norway stationary	27	68.12	14.44	March/2003
North Sea	40	58	-3	March/2003

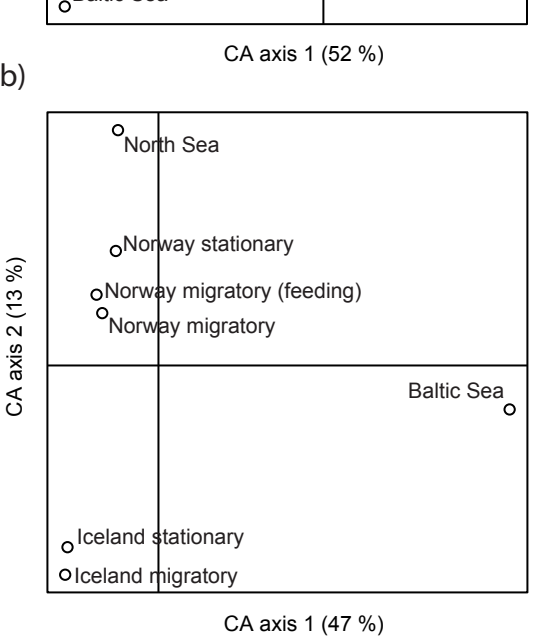
Baltic Sea	40	54.87	15.46	April/1997

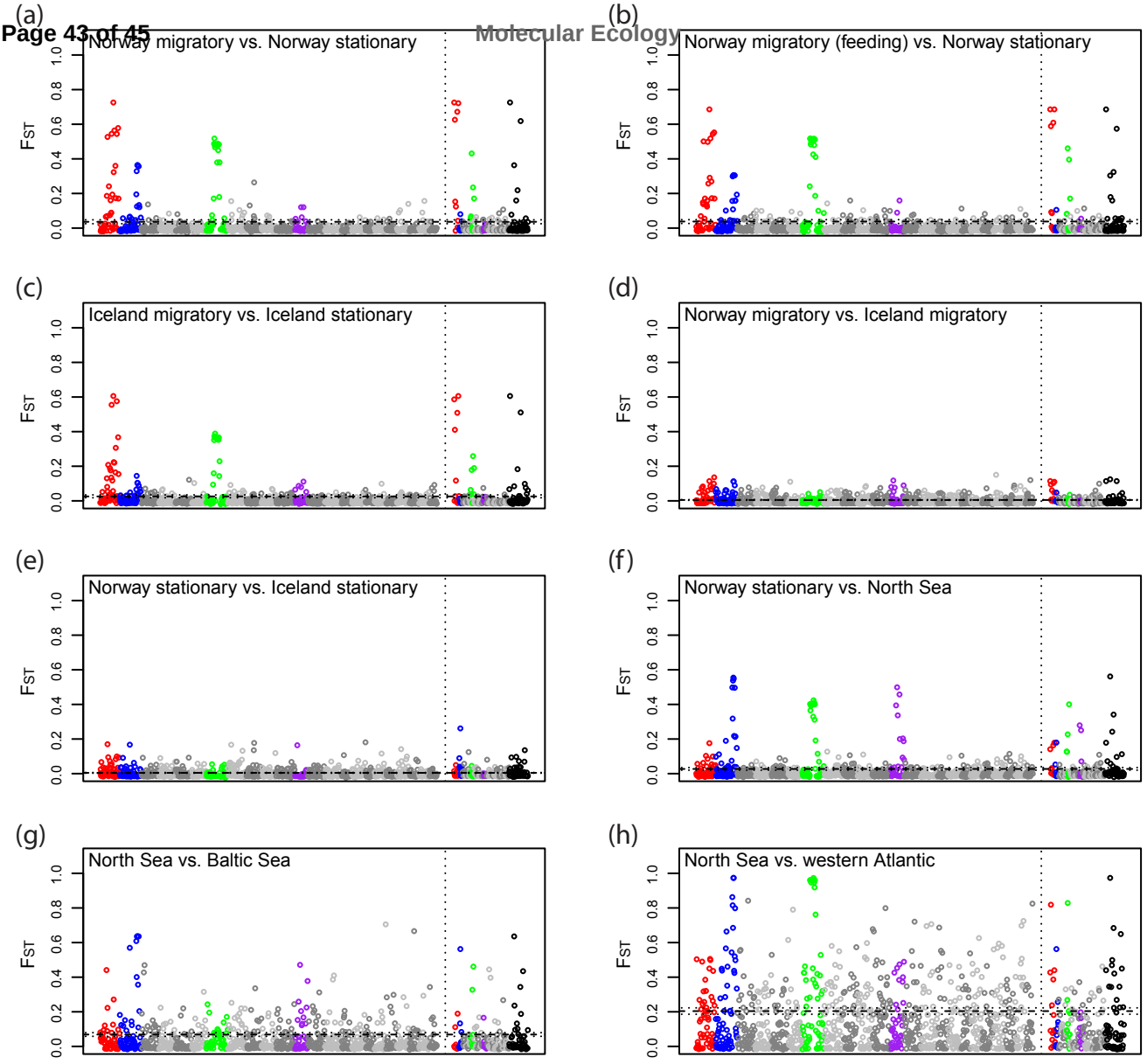

Norway migratory (feeding)
Western Atlantic
Norway migratory
Norway stationary
Iceland stationary
Iceland migratory
North Sea
Baltic Sea

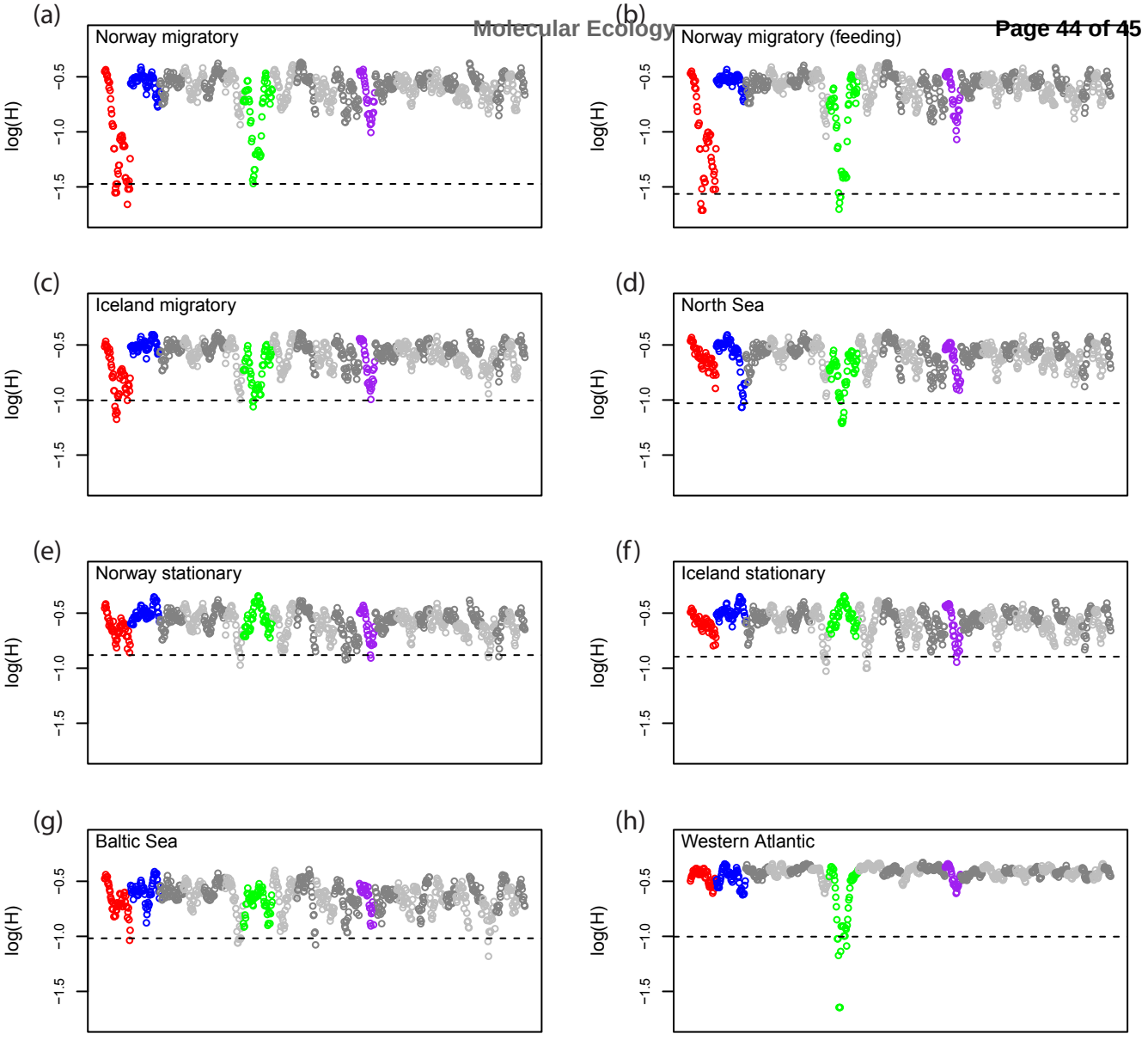
(a)

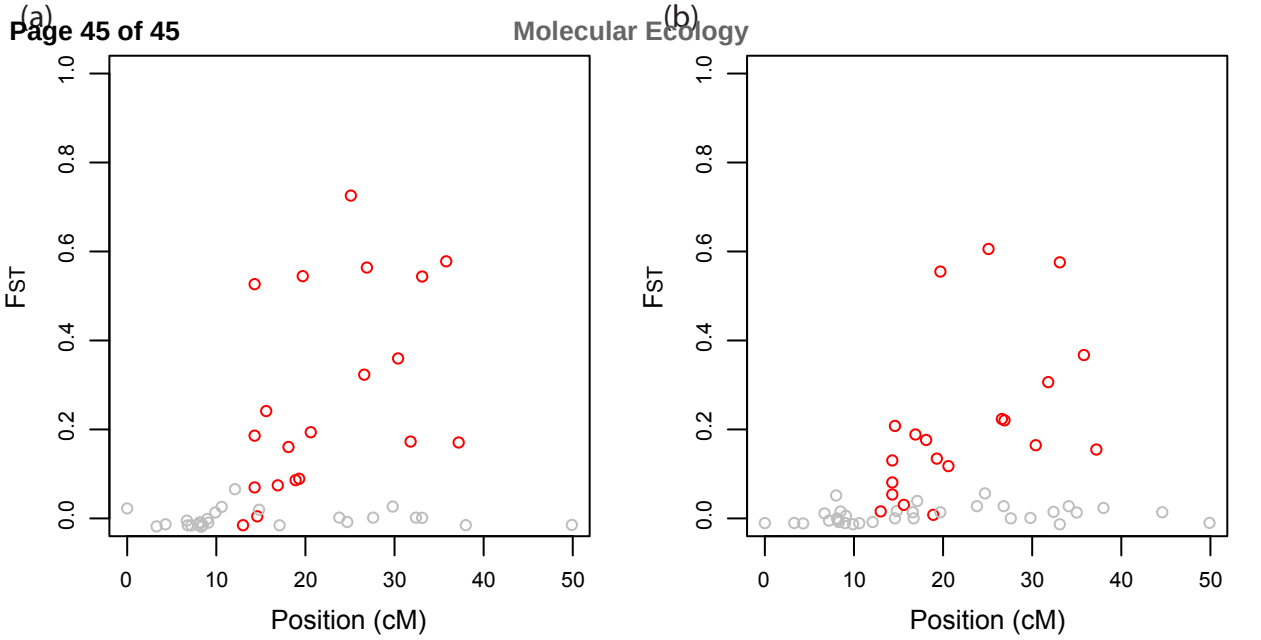

North Sea
Iceland stationary
Norway stationary
Iceland migratory
Norway migratory (feeding)
Norway migratory
Baltic Sea
CA axis 2 (23 %)
CA axis 1 (52 %)

(b)


North Sea
Norway stationary
Norway migratory (feeding)
Norway migratory
Baltic Sea
Iceland stationary
Iceland migratory
CA axis 2 (13 %)
CA axis 1 (47 %)


(a)
Norway migratory vs. Norway stationary
(b)
Norway migratory (feeding) vs. Norway stationary
(c)
Iceland migratory vs. Iceland stationary
(d)
Norway migratory vs. Iceland migratory
(e)
Norway stationary vs. Iceland stationary
(f)
Norway stationary vs. North Sea
(g)
North Sea vs. Baltic Sea
(h)
North Sea vs. western Atlantic
FST
0.0
0.2
0.4
0.6
0.8
1.0


(a)
Norway migratory
(b)
Norway migratory (feeding)
(c)
Iceland migratory
(d)
North Sea
(e)
Norway stationary
(f)
Iceland stationary
(g)
Baltic Sea
(h)
Western Atlantic
log(H)
−0.5
−1.0
−1.5


(a)
(b)
FST
Position (cM)
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
40
50